STAT 201B: Probability

Feynman Liang*
Department of Statistics, UC Berkeley

Last updated: January 28, 2020

Contents

1 Lecture 1: A high level overview

2020-01-21

Instructor Jim Pitman, Evans 303

Text Durrett's PTE [durrett2019probability Dur19], Version Jan 2019[1]

Topics

- Mainly chapters 5–9 of Durrett's
- Material from chapter 3 neglected in 205A
- Poisson processes
- Infinitely divisible and stable laws

We will start with Markov chains; read Chapter 5 of Durrett in next few weeks.

2 Lecture 2: Markov chains

2020-01-22

2.1 Kolmogorov consistency

A crucial tool for proving existence of sectionchastic processes is *Kolmogorov's consistency Theorem.* Suppose we want to make a stochastic process

$$(X_i, i \in I), \quad X_i \in \mathbb{R}$$

perhaps using a large index set I.

Observe that for every $F \subset I$, $|F| < \infty$, the law of $(X_i, i \in I)$, presuming it exists on $\prod_I \mathbb{R}$, induces a joint law on $(X_i, i \in F)$.

$\Pr(\wedge_{j=1}^n X_{i_j} \le x_i)$ is known from $\prod_I \mathbb{R}$, and this is the CDF of something on $\mathbb{R}^n$. Hence, the law of $(X_i, i \in F)$ is a probability measure on $\mathbb{R}^{|F|}$. The law on $\mathbb{R}^{|F|}$ and the law on $\prod_I \mathbb{R}$ are related: *they must be consistent*! A joint law on (X_1, X_2) implies laws for X_1 and X_2 obtained by projection (i.e. marginalization).

More generally, if $F \subset G$ then the law of $X_F := (X_i, i \in F)$ must be the projection of the law of X_G.

*feynman@berkeley.edu
[1] https://services.math.duke.edu/~rtd/PTE/PTE5_011119.pdf

Definition 1

A collection of joint laws of X_F on $\mathbb{R}^{|F|}$ for each $F \subset I$ finite is *consistent* if it satisfies the above.

Theorem 2 (*Kolmogorov*)

For every consistent collection of finite dimensional distributions (FDDs), there exists a unique probability law on $(\mathbb{R}^I, \otimes_{i\in I}\mathcal{B})$ with these prescribed FDDs.

Remark 3. We hardly ever use it, because in nearly all interesting cases it is obvious how to construct $(X_i, i \in I)$ once you know the FDDs.

Remark 4. Unless I is countable, there is a serious issue that too few sets in $\mathbb{R}^I$ are measurable for us to care about this product construction.

Exercise 5. Show that every measurable set in $\mathbb{R}^I$ is generated by some countable set of coordinates: $B \subset \mathbb{R}^I$ in product σ-field, then $B \in (X_{i_1}, X_{i_2}, \dots)$ for some sequence $i_1, i_2, \dots$.

As a result, with $I = [0, 1]$ the set of continuous paths is not measurable!

Underfull vbox

Example 6 (*Continuum of IID r.v.s*)

Given any consistent family of FDDs, we can construct an infinite sequence of IIDs (X_i) with any prescribed law F.
Let $\omega \in [0, 1] \to (X_1(\omega), X_2(\omega), \dots)$ be the binary expansion of $\omega = \sum_{n=1}^{\infty} X_n(\omega)2^{-n}$.
We can construct infintely many IID uniforms on $[0, 1]$ by setting

$$U_1 = \sum_{i\geq 1} X_{2^i} 2^{-i}$$
$$U_2 = \sum_{i\geq 1} X_{3^i} 2^{-i}$$

continuing with other powers of primes to maintain independence. Then use inverse CDF to construct arbitrary law.

2.2 Markov chains

{sub:Markov␣chains}

Let $(S, \mathcal{S})$ be the *state space*, sometimes abstract. If S is countable, we will always take $\mathcal{S} = 2^S$.

Good to know what works easily for $(S, \mathcal{S})$.

Start with a general definition.

Definition 7 (*Markov transition function*)

Let $(S_1, \mathcal{S}_1)$ and $(S_2, \mathcal{S}_2)$ be two probability spaces. P is called a *Markov transition function / kernel* if

$$P : S_1 \times \mathcal{S}_2 \to [0, 1]$$

satisfying

1. For all $x \in S_1$, $A \mapsto P(x, A)$ is a probability measure on $(S_2, \mathcal{S}_2)$.
2. For fixed $A \in \mathcal{S}_2$, $x \mapsto P(x, A)$ is $\mathcal{S}_1$-measurable.

Usually $\mathcal{S}_2 = \sigma(\mathcal{C}_2)$ for some π-system $\mathcal{C}_2$ (closed under intersection), so we only need to check measurability of $x \mapsto P(x, A)$ for $A \in \mathcal{C}_2$. This is because (assuming 1.) $\{A : x \mapsto P(x, A) \text{ measurable}\}$ is a λ-system containing π-system $\mathcal{C}_2$, apply Dynkin π-λ.

Theorem 8

Let λ be a probability distribution on $(S_1, \mathcal{S}_1)$ and $P(\cdot, \cdot)$ a transition function for $(S_1, \mathcal{S}_1) \to (S_2, \mathcal{S}_2)$. Then $\exists!$ joint law of (X_1, X_2) on $(S_1 \times S_2, \mathcal{S}_1 \otimes \mathcal{S}_2)$ such that

$$X_1 \sim \lambda$$
$$X_2 \mid X_1 = x \sim P(x, \cdot)$$

{thm:fubini}

Theorem 9 (*Fubini*)

$$\mathbb{E}g(X_1, X_2) = \int_{S_1}\int_{S_2} g(x_1,x_2) \underbrace{\Pr(X_1 \in dx_1)}_{\lambda(dx_1)} \underbrace{\Pr(X_2 \in dx_2 \mid X_1 = x_1)}_{P(x_1, dx_2)}$$

for every product measurable g which is either bounded, non-negative, or absolutely integrable.

Exercise 10. $\mathbb{E}[g(X_1, X_2) \mid X_1] = \psi_g(X_1)$ where $\psi_g(x) =$??

Proof. Key idea: make the joint law on $(S_1 \times S_2, \mathcal{S}_1 \times \mathcal{S}_2)$ first. Construct measure on product space, X_1 and X_2 are the coordinates.

Define

$$I(g) = \int_{S_1} \lambda(dx_1) \int_{S_2} g(x_1, x_2) P(x_1, dx_2)$$

Potential issues with $I(g)$:

- Is it defined? Try $g(x_1, x_2) = f(x_1)h(x_2)$ for $f \in m(\mathcal{S}_1)$ and $h \in m(\mathcal{S}_2)$.
- Must check the second integral is $\mathcal{S}_1$-measurable
- OK because take $f = \mathbb{1}_A$, $h = \mathbb{1}_B$.

$g \mapsto I(g)$ is linear, bounded, non-negative, monotone. What is it's measure? Take $g = \mathbb{1}_C$ for $C \in \mathcal{S}_1 \otimes \mathcal{S}_2$ and use monotonicity of the two integrals in I along with monotone convergence theorem.

Hence, $C \mapsto I(\mathbb{1}_C)$ is a probability measure on the product space satisfying the desired goals. □

Remark 11. This is only abstract measure theory. If we were in say $\mathbb{R}^2$, we could appeal to Riesz representation theorem.

How?

Flipping the transition kernel

There is no universal Bayes theorem for inverting laws (X_1), CD of $(X_2 \mid X_1) \leftrightarrow$ law (X_2), CD of $(X_1 \mid X_2)$

$$\Pr[X_2 \in A] = \int \Pr[X_1 \in dx_1] \underbrace{\Pr(X_2 \in A \mid X_1 = x)}_{P(x,A)}$$

Only if we assume densities relative to a product measure, a great ingenuity can gen flip general Markovian P to get a $\tilde{P}$.
In the discrete case, these are a matrix and vector:

$$P = (P(x, \{y\}), x \in S_1, y \in S_2) \tag{1}$$
$$\lambda = (\lambda(x), x \in S) \tag{2}$$

$$\lambda(x) = \Pr(X_1 = x) \tag{3}$$
$$\Pr(X_2 = y) = \sum_x \lambda(x) P(x,y) = (\lambda P)(y) \tag{4}$$
$$\Pr(X_1 = x \mid X_2 = y) = \frac{\lambda(x,y)P(x,y)}{(\lambda P)(y)} \tag{5}$$

for all $y : (\lambda P)(y) > 0$.

More generally, let λ be the initial distribution of X_1, P a transition kernel $X_1 \to X_2$, and $f \geq 0$

nonnegative measurable.

$$\mathbb{E}f(X_2) = \int_{S_1} \lambda(dx)(Pf)(x) = \int_{S_2} (\lambda P)(dy) f(y)$$
$$Pf(x) = \int_{S_2} P(x, dy) f(y)$$
$$(\lambda P)(A) = \int_{S_1} \lambda(dx) P(x, A)$$

Notice $\lambda \mapsto \lambda P$ and $f \mapsto Pf$ have excellent and obvious properties as operators on probability measures.

Example 12

If $f \geq 0$, then $Pf \geq 0$.
If $0 \leq f_n \uparrow f$, then $0 \leq Pf_n \uparrow Pf$.

If S_1 and S_2 are finite, then we are doing matrix algebra whree P is a $S_1 \times S_2$ stochastic matrix and λP is the action of P on a row vector λ (distributions over states, measures more generally) and Pf the action on a column vector f (observables over states, non-negative functions more generally).

Now consider $S_1 = S_2$ countable.

Definition 13

Say (X_n) is a Markov chain with initial distribution λ and stationary transition matrix P if

$$\Pr[X_0 = x_0, \ldots, X_n = x_n] = \lambda(x_0) \prod_{i=1}^{n} P(x_{i-1}, x_i)$$

3 Lecture 3: More on Markov chains

2020-01-28

A few more things about Markov chains which we can do with little cost, with general state space $(S, \mathcal{S})$, discrete time $n = 0, 1, \ldots$.

General view of how to construct a sequence of RVs $(X_i)_{i \geq 0}$:

1. Make $X_0 \sim \lambda$, $\Pr(X_0 \in A) = \lambda(A)$ where $\lambda = \delta_x$ = Dirac unit mass at x.
2. Given $X_0 = x$, we have a Markov transition kernel
$$P(x_0, \cdot) = \Pr(X_1 \in \cdot \mid X_0 = x_0)$$
Construct (X_0, X_1) using Fubini
$$\mathbb{E}g(X_0, X_1) = \int_S \lambda(dx_0) \int_S P(x_0, dx_1) g(x_0, x_1)$$
3. Given $X_0 = x_0$ and $X_1 = x_1$, make X_2 according to
$$P(x_0, x_1, \cdot) = \Pr(X_2 \in \cdot \mid X_0 = x_0, X_1 = x_1)$$
We can do that! Nothing new required since
$$X_0 \mapsto (X_0, X_1), \qquad \lambda \mapsto \lambda \otimes P \text{ on } S \times S$$
Now
$$\mathbb{E}g(X_0, X_1, X_2) = \underbrace{\int_S \lambda(dx_0) \int_S P(x_0, dx_1)}_{\int_{S \times S} (\lambda \otimes P)(dx_0, dx_1)} \int_S P(x_0, x_1, dx_2) g(x_0, x_1, x_2)$$

if S is nice, every stochastic process $(X_i)_{i \geq 0}$ is "like" (i.e. FDDs are equal in distribution) the above construction.

Definition 14

Such a process is called *Markov* iff it is equal in distribution to a process made with a kernel

$$P(dx_{n+1} \mid x_0, \dots, x_n) = P_n(dx_{n+1} \mid x_n)$$

If $P_n \equiv P$, then the process is called *homogeneous*.

- Not mmuch general theory of inhomogeneous
- We will always henceforth assume homogeneous
- Mostly, discrete S
- Techniques for general S
  - Figure it out for finite/countable S
  - Write the args so they wrook more generally
- Note: if you upgrade $X_n = Y_n := (X_0, \dots, X_n)$
- Technical point: can only use Fubini to prescribe FDDs. From our schemme, we create consistent FDDs and are left with two options:
  1. Make the chain explicitly as $X_n =$ some measurable function on $(\Omega, \mathcal{F}, \Pr)$. For example, for IID uniforms take $X_n = F^{-1}(U_n)$ for an IID sequence U_n. Then you have RVs $(X_i)_{i \geq 0}$ on $(\Omega, \mathcal{F}, \Pr)$ which implies a law on $\prod_{i \geq 0} S$ given by

     $$\omega \mapsto (X_0(\omega), X_1(\omega), \dots)$$

  2. If not, we need Kolmogorov consistency or better (Ionescu-Tulcea Theorem, which says the infinite process exists with no regularity on $(S, \mathcal{S})$).

Fix $(S, \mathcal{S})$ and Markov kernel P. For $f \geq 0$ bounded measurable

$$(Pf)(x) := \int_S P(x, \cdot) f(\cdot) = \mathbb{E}[f(X_1) \mid X_0 = x]$$

Convention: We write $\mathbb{E}[f(X_1) \mid X_0 = x] = \mathbb{E}_x f(X_1)$, or more generally $\mathbb{E}[f(X_1) \mid X_0 \sim \lambda] = \mathbb{E}_\lambda f(X_1)$.

Example 15

$\mathbb{E}_\lambda f(X_0, X_1, X_2) = \int_S \lambda(dx) \mathbb{E}_x f(x, X_1, X_2)$ by the Fubini theorem. This is an example with a regular conditional distribution (RCD) for $(X_0, X_1, X_2 \mid X_0 = x)$.
More generally, anytime we see an expectation involving an integral we are really just doing conditioning. Most things in Markov chains are obvious by writing down Fubini.

A special case: $\mathbb{E}_x f(X_1) = f(x)$ for all $x \in S$ means that $(P - I)f = 0$ which means that $(f(X_0), f(X_1))$ is a martingale pair relative to $\sigma(X_0)$, $\sigma(X_0, X_1)$. More generally, for any initial distribution λ the collection $(f(X_i))_{i \geq 0}$ is a martingale relative to $\mathcal{F}_n = \sigma(X_i : i \leq n)$. For a martingale

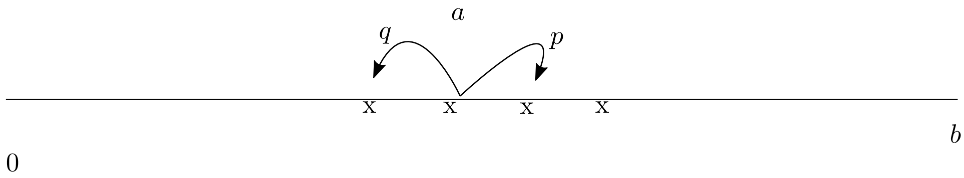
$$\mathbb{E}[f(X_{n+1} \mid X_0, \dots, X_n] = g(X_n)$$

by the time homogeneous markov property, we must have $g = Pf$. So if $Pf = f$, then $\mathbb{E}[\cdot] = f(X_n)$. This is the definition of a MG.

Example 16 (*Gambler's Ruin*)

{eg:gamblers-ruin}

$0 \leq a \leq b$, Gambler starts with a dollars and at each turn moves $+1$ with probability p and -1 with probability q. Continue until reach 0 or b and then stop.



{fig:gamblers-ruin}

Figure 1: Gambler's ruin

We can write this as $S_0 = a$ and $S_n = a + \sum_i X_i$ where X_i are ± 1 with probabilities p/q. Define (the stopping time)

$$T_{0b} = \min\{n \geq 1 : S_n \in \{0, b\}\}$$

Our MC is then $(S_{n \wedge T_{0b}} : n \geq 0)$. Viewed as a MC, we have an absorbing chain and are interested in $\Pr_a[\text{absorbed at } b]$.
Alternatively, viewed as a RW we are interested in

$$\Pr_a[S_{T_{0b}} = b]$$

By Wald's identity

$$\mathbb{E}[S_{T_{0b}} - a] = \mathbb{E}[X]\mathbb{E}[T_{0b}]$$

For fair case, $p = q = 1/2$ and $\mathbb{E}X = 0$.

$$(b-a)\Pr_a[\text{absorb at b}] + (0-a)\Pr_a[\text{absorb at 0}] + \underbrace{??}_{0} = 0$$

$$\Pr_a[\text{absorb at b}] + \Pr_a[\text{absorb at 0}] = 1$$

How do we know it eventually hits and doesn't oscillate forever? Bounded by Geometric RV, so goiing to be finite with probability 1 (could have use martingale convergence, but dont need).
In the unfair case, we need a better martingale (Wald MG). Try an exponential martingale

$$Z^{S_n} \text{ for suitable } Z$$

To find Z, we would like

$$\begin{aligned}
\mathbb{E}[Z^{\overbrace{=S_n+X_{n+1}}^{S_{n+1}}} \mid S_0, \ldots, S_n] &= Z^{S_n} \\
\Leftrightarrow \mathbb{E}Z^{X_{n+1}} &= 1 \\
\Rightarrow Z &= q/p
\end{aligned}$$

Applying Wald's identity again, we solve the system

$$\begin{aligned}
\left(\frac{q}{p}\right)^a &= \Pr_a[\text{abs at b}](q/p)^b + \Pr_a[\text{abs at 0}](q/p)^0 \\
1 &= \Pr_a[\cdot] + \Pr_a[\cdot]
\end{aligned}$$

Within the Markov chain frammework, we say that we seek a harmonic function h such that $h = Ph$ and $h(0) = 0$, $h(b) = 1$. Then

$$\Pr_a[\text{hit b before 0}] = ??$$

Notice $h_0(x) = (q/p)^x$ and $h_1(x) \equiv 1$ both solve $h = Ph$, so finding the linear combination that satisfies boundary conditions yields

$$h(x) = \frac{(q/p)^x - (q/p)^0}{(q/p)^b - (q/p)^0}$$

Theorem 17

Suppose you have a MC with absorbing states (i.e. boundary states) $B \subset S$ and target states $A \subset B$.

$$\Pr_a[X_n \in A \text{ for all large } n]$$

Let

$$h_A(x) = \Pr_x[\text{absorved in } A]$$

Argue $h(x) = h_A(x)$ must solve $h = Ph$ using a one-step analysis:

$$h(x) = \sum_{y \in S} P(x,y)h(y)$$

by conditioning on $X_1 = y$
Notice $h_A(x) = \lim_n P^n \mathbb{1}_A$ and

$$(P^n \mathbb{1}_A)(x) = \Pr_x[X_n \in A]$$

Since A is absorbing, $X_n \in A \Rightarrow X_{n+1} \in A$ so $\{X_n \in A\}$ is increasing, so $h_A = Ph_A$.
Also, $h_A = \mathbb{1}_A$ on B because B is absorbing.

Since Dirichlet BVP has unique soution, we hope/expect that there is a unique solution of $h = Ph$, $h = \mathbb{1}_A$ on B, and that in this case $h = h_A$.

References

{durrett2019 probability} [Dur19] Rick Durrett. *Probability: theory and examples.* Vol. 49. Cambridge university press, 2019.